The Charles Phelps Taft Research Center and
the Department of Mathematical Sciences
present

Alan F. Karr

Director, National Institute of Statistical Science

What is Data Confidentiality and Why Should I Care?

Room 800 Swift Hall
Tuesday, May 8, 2007 4 – 5 pm

Many organizations, among them government agencies, corporations, employers, and medical insurers, assemble and, in many cases, disseminate highly-sensitive information about individuals and establishments. Some do so responsibly, while others are less careful. The use of the internet as a mode of communication exacerbates problems, but also offers paths to solution.

Data confidentiality is the field within the statistical and computer sciences that addresses how to make trade-offs in today's electronic world between the conflicting goals of protecting the privacy of data subjects and making useful information available from the data for policy and research purposes.

This talk will be an accessible introduction to data confidentiality research, emphasizing the nature and scale of the issues, methods by which data can be made safe for dissemination, and current research problems. Disclosure risk and data utility serve as framing abstractions: data disseminators should make decisions based on quantified measures of risk and utility.

Secure Statistical Analysis of Distributed Databases

Room 500 Swift Hall
Wednesday, May 9, 2007 10 – 11 am

A continuing need - in the contexts from national security to business - is for statistical analyses that integrate data stored in multiple, distributed databases. But, barriers to integrating databases, such as confidentiality and scale, are numerous, and often literal integration is impossible.

In this talk, we show how many analyses can be performed without integrating the data. Using techniques from computer science known as secure multi-party computation, the database holders can share analysis-specific sufficient statistics anonymously, but in a way that the desired analysis can be performed in a statistically valid manner. Several illustrative analyses will be presented: secure regression for both horizontally and vertically partitioned data; secure data integration; secure contingency tables; and secure maximum likelihood estimation.

Simple implementations of the protocols are subject to a variety of threats, including corruption by outsiders, collusion among database holders, and unilateral incentives for holders to "cheat" by reporting false data or sufficient statistics. We discuss ways of reducing the threats. In particular, partially trusted third parties (PTTPs), which hold some information not available to the database holders, will be described.

For more information, please contact Professor Xiaodong Lin at linxd@math.uc.edu